

*Aligning the NWEA RIT Scale with the
Pennsylvania System of School Assessment (PSSA)*

January, 2004

John Cronin, Ph.D. and Branin Bowe – Northwest Evaluation Association

Each year, Pennsylvania students participate in testing as part of the Pennsylvania assessment program. Students in grades 5, 8, and 11 take tests in reading and math while those in grades 6, 9 and 11 are assessed in writing. These tests serve as an important measure of student achievement for the state's accountability system. Results from these assessments are used to make state-level decisions concerning education, to meet *Adequate Yearly Progress* (AYP) reporting requirements of the *No Child Left Behind Act* (NCLB), and to inform schools and school districts of their performance.

The Pennsylvania Department of Education has developed scales that are used to assign students to one of four performance levels on the state's assessments. These are, from the lowest cut score to the highest: *below basic*, *basic*, *proficient*, and *advanced*. For purposes of NCLB, the *proficient* level is considered the level that represents satisfactory performance.

Many students who attend school in Pennsylvania also take tests developed in cooperation with the Northwest Evaluation Association (NWEA). These tests report student performance on a single, cross-grade scale, which NWEA calls the RIT scale. This scale was developed using Rasch scaling methodologies. RIT-based tests are used to inform a variety of educational decisions at the district, school, and classroom level. They are also used to monitor academic growth of students and cohorts. Districts choose whether to include these assessments in their local assessment programs. They are not state mandated.

The versions of NWEA tests in use in Pennsylvania have been specifically aligned to match the content of local and Pennsylvania state curriculum standards. Because of this, we believe there is a good match in content between the NWEA tests and the curriculum standards being used in Pennsylvania.

In order to use the two testing systems to support each other, an alignment of the scores from the state and RIT-based tests is as important as the curriculum alignment. The current study is one of an ongoing series of studies that are being conducted to identify the relationships between NWEA tests and state-mandated assessments. Studies in sixteen states have now been completed. For purposes of this study we focused on examining the relationships between PSSA and NWEA assessments in reading and mathematics only.

The primary questions addressed in this study are:

- To what extent do the same subject scores for the NWEA test correlate to the content-similar subjects on the PSSA tests?
- What RIT scores correspond to various performance levels on the PSSA tests?
- How well can *proficient* performance on the Pennsylvania assessments (be predicted from RIT scores when NWEA assessments are administered in the same time frame?

Method

Participating School Systems

An e-mail solicitation was sent in October, 2003 to all Pennsylvania school systems who had two or more seasons of experience with NWEA testing prior to spring 2003 to secure participants for the study. Based on the response from this solicitation, spring 2003 PSSA and NWEA student assessment records in reading and mathematics were collected from two school districts.

Data Preparation

For purposes of studying NWEA test alignment with the PSSA, 5th and 8th grade student level test records from spring 2003 PSSA testing and spring 2003 NWEA assessments were matched using district assigned student ID numbers. Matched records were then screened to remove invalid scores. Table I shows the number of student records included in the reading analysis for this study.

Table 1
Reading and Mathematics Tests Included by Grade

Subject	5	8	Total
Reading	1365	1075	2440
Mathematics	1365	1075	2440

We had enough student records at each grade to adequately cover the breadth of the scale and perform a robust analysis near the proficiency point for both grades in reading and math. In 8th grade there were a relatively small number of students (44 in math and 78 in reading) who performed at the *advanced* level. Although we believe our final estimates of the cut score for *advanced* performance in reading and math are reasonable estimates based on the data, they may not be as robust as our estimates for the other performance levels.

Because the study involved a small number of districts, we recommend that schools validate our estimates by cross-checking their own students' performance against our cut scores.

Analyses

Pearson correlations. The initial analyses focused on the relationships among the NWEA and Pennsylvania assessment scores at each grade to determine how closely the scores on the NWEA test correlated with same subject scores on the PSSA. Simple bivariate correlation coefficients were computed among these scores.

Linking PSSA scores to the RIT scales. Three methods of estimating cut scores for PSSA levels were used. The most straightforward was simple linear regression ($PSSA_{pred} = a(RIT) + c$). Since we sometimes observe departures from a linear relationship on the lower and upper ends of state test scales, a second order regression model was also used ($PSSA_{pred} = a(RIT^2) + b(RIT) + c$). For each of these methods, the RIT score was determined by substituting the appropriate PSSA score for $PSSA_{pred}$ and solving the equation for RIT.

A fixed-parameter Rasch model was also used to estimate RIT cut scores. In this method, the PSSA performance level was treated as a test item. The assumption is that the performance level 'item' should contain all the information about the difficulty of the test. Student abilities (RIT

scores) were the 'fixed parameter' used to anchor the difficulty estimate of the 'status' item to the RIT scale. The resulting 'difficulty estimate' was taken as the RIT cut score for this method. This is referred to as the Rasch Status on Standard (or simply Rasch SOS) method.

Predicting PSSA performance levels from RIT scores. RIT scores were first used to predict whether students were likely to achieve performance at or above the *proficient* performance level on the PSSA. We make the estimates from this level in order to maintain consistency with prior studies of state test alignment, which make comparisons based on the NCLB reported performance level. This allows us to make accurate comparisons of our alignment with different state tests.

The predictions of PSSA performance were compared to observed performance in 2 X 2 contingency tables. A *prediction index* score was generated to measure the ratio of Type I error to accurate prediction of proficiency status. This score is expressed as

$$1-(\text{Number of Type I errors}/\text{Number of correct predictions})$$

Higher prediction index numbers generally show more accurate prediction with lower levels of Type I error. Type I error occurs when NWEA assessments predict that a student will achieve above a passing level of performance when the student actually achieves a failing score. This index was generated for the linear, second order, and Rasch SOS methodologies. In general, the highest prediction index score was used to select the RIT cut score to be adapted as the *official* RIT score we would associate with achieving the passing standard on the corresponding PSSA assessment for the particular grade level and subject area. We do make exceptions to this rule when the estimated score produces high accuracy rates but inordinately large numbers of Type II errors. This condition indicates a greatly overestimated cut score, so we select a method that produces a more balanced Type I to Type II error ratio in these instances.

In addition, we evaluated the accuracy of predictions of PSSA levels based on observed RIT scores. The predictions of PSSA level performance were compared to observed performance in 4 X 4 contingency tables. Once again a prediction index score was generated to provide an estimate of accuracy.

Content Validity

Formal comparisons of the content of NWEA and Pennsylvania tests were not conducted for purposes of this study. The standards used to construct the NWEA Assessments were the same as those used for the Pennsylvania assessments. Both NWEA assessments and the Pennsylvania assessments include multiple-choice items. The PSSA also includes short answer and extended response questions. Results from our previous fifteen studies indicate that the addition of items in alternate formats generally does not, by itself, materially affect the ability of the NWEA test to generate reasonably accurate predictions of performance levels.

Results

Descriptive Statistics

Table 2 reviews descriptive statistics for the PSSA and NWEA assessments. The median RIT scores for this sample are far below those for the NWEA norm population. In reading, the median score for the sample is 4 points below the median of our norm population in grade 5 and 11 points below the national norm median in grade 8. In mathematics, the mean score for the sample is 9 points below the national norm median in grade 5 and 13 points below in grade 8.

These differences are large and their potential impact on the accuracy of our estimates merits discussion.

Normal distributions around a nationally-normed mean are desirable but not necessarily essential when conducting alignment studies. It is more important that the sample provide reasonable numbers who perform at all levels on the test scales than normal distribution so that the statistical methods applied have an adequately large sample to derive good estimates of performance levels that are at the higher and lower ends of a test scale. With the exception already noted for the advanced level in 8th grade mathematics, we had reasonably large representations of students who performed at all performance levels.

It is fair to say, however, that school districts with large numbers of low performing students may align their curriculum differently to the state standards. There may also be other, hard to know factors, related to this phenomenon that may influence alignment. That's why we recommend that school systems test the application of the study results in their own setting to validate the predicted cut score's accuracy.

Table 2
Means, Standard Deviations, and Medians for the PSSA and NWEA assessments

Grade	5	8
PSSA Reading		
N	1365	1075
Mean	1259.80	1185.62
Median	1250	1198
Std. Deviation	229.28	212.54
NWEA Reading		
N	1365	1075
Mean	206.40	212.53
Median	208	214
Std. Deviation	16.87	17.41
PSSA Mathematics		
N	1365	1075
Mean	1308.69	1192.55
Median	1299	1171
Std. Deviation	204.13	154.20
NWEA Mathematics		
N	1365	1075
Mean	215.93	221.25
Median	215	222
Std. Deviation	15.77	17.52

Pearson correlations

Table 3 shows the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and PSSA. Same subject correlations were very high, ranging from .84 to .87, numbers that suggest the tests were generally measuring the same constructs. Discriminant validity was tested by examining same subject Pearson correlations next to correlations for the alternate subject (math against reading). In all cases the same subject correlations were higher than correlations against the alternate subject.

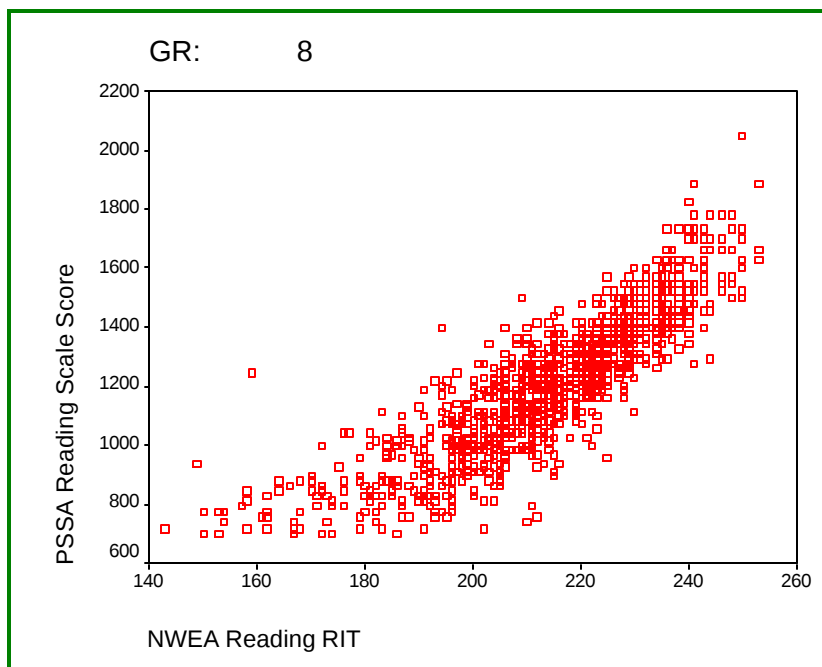
Table 3
Pearson Correlations for PSSA and NWEA assessments by Subject

Grade 5 (n=1365)				
	PSSA Reading	NWEA Reading	PSSA Math	NWEA Math
PSSA Reading	1	.844	.794	.797
NWEA Reading		1	.726	.797
PSSA Math			1	.872
NWEA Math				1
Grade 8 (n=1075)				
	PSSA Reading	NWEA Reading	PSSA Math	NWEA Math
PSSA Reading	1	.844	.738	.771
NWEA Reading		1	.691	.795
PSSA Math			1	.848
NWEA Math				1

- Same subject correlations are shaded

Analysis of scatterplots suggested that relationships might be slightly curvilinear, and that some of the scale relationships might break down slightly near the lower end of the scales, possibly indicating a floor effect on the PSSA. Figure 1 provides an example from the 8th grade reading sample that illustrates both the scale relationships and the evidence of some breakdown in correlation near the bottom of the PSSA Scale. For example, note that students achieving scores near 700 on the PSSA scale, achieve scale scores between 140 and 210 on the NWEA test. One possible explanation for this is that the NWEA test, because it is adaptive as opposed to single form, has the capacity to more accurately measure performance at the low end of performance.

Figure 1 – Scatterplot depicting Grade 8 NWEA reading RIT against the Grade 8 PSSA reading scale score



Linking PSSA performance level cut scores to the RIT scale

The primary purpose of this study was to estimate the RIT scale scores that most closely correspond to the cut scores for different performance levels on the PSSA. This information allows schools to identify students who may need additional support to reach state standards. It can also help schools identify students who are performing well enough that they are ready to tackle work beyond what the state standards require.

Table 4 shows several estimations of the Spring 2003 RIT score that correspond to the cut scores for the various performance levels on the PSSA scales. As a rule the three methodologies came to very similar estimates of cut scores for each of the performance levels. As expected, the estimates for the advanced cut score in grade 8 mathematics varied greatly, primarily because the estimates are based on very small numbers of students who actually performed at that level on the PSSA.

Table 4
Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the PSSA

	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Reading	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 5	<=198	199	210	226	<=200	201	212	224	<=200	201	210	221
Grade 8	<=207	208	222	242	<=210	211	223	238	<=209	210	221	235
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Mathematics	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 5	<=204	205	216	231	<=204	205	216	231	<=205	206	216	228
Grade 8	<=219	220	237	264	<=222	223	237	254	<=221	222	236	250

Predicting PSSA pass-fail status from RIT scores

Once the cut scores were estimated from the three methods, we evaluated each possible cut score to determine how accurately it predicted students' actual performance on the corresponding PSSA assessment. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different PSSA performance levels. A *prediction index* statistic (described on page 3) scored the accuracy of prediction.

For this study, we first assessed the accuracy of the RIT scale in correctly predicting whether students are likely to reach the *proficient* level on the corresponding PSSA test. Next we assessed the accuracy with which the RIT predicted level assignment on this test. Use of the prediction index statistic helped assure that the method chosen produced a high ratio of accurate passing predictions relative to Type I errors. Type I errors occur when the RIT scale predicts a passing score for a student who actually fails the assessment. These types of errors raise particular concern because they fail to identify students who might need additional support and resources in order to achieve their targets. A high prediction index number indicates that the test maximizes accuracy of prediction while minimizing Type I errors.

In these kinds of studies we want to emphasize that prediction is not used to foretell an inevitable future for the student, rather it is used to help schools plan for instruction and offer appropriate interventions to children who need additional support to be successful. For purposes of the *No*

Child Left Behind Act, schools are judged on their ability to move children to the *proficient* level and beyond. RIT scores can provide teachers with advance notice about students who may not reach these goals on the Pennsylvania assessment that corresponds to their grade level.

Table 5 shows the results for reading. All methods considered were highly accurate (better than 84%) in predicting pass-fail against the proficiency cut score. The second-order methods generated fewer Type I errors in both grades (about 5 to 6%). All methods produced prediction index scores above .890. The results suggest that the NWEA reading assessments predicted proficiency status on the corresponding PSSA assessments very well.

Table 6 shows the results for mathematics. Once again all methods considered were quite accurate; each generated accuracy rates above 84% for 5th grade and 89% for 8th grade. All methods employed limited Type I errors to below 9% of cases in grade 5 mathematics and to below 5% of cases in grade 8 mathematics.

The Rasch method of calibration generally produced more accurate predictions of proficiency status than either regression method in all subjects. Linear and second order regression produced identical statistics and were slightly more accurate than Rasch SOS methods with this sample population.

Table 5
Accuracy of the RIT scale in predicting PSSA proficiency status – reading

Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	210	84.38%	9.16%	.891
Second Order	212	85.26%	6.16%	.928*
Rasch SOS	212	85.26%	6.16%	.928*
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	222	85.86%	5.86%	.932
Second Order	223	85.86%	4.84%	.944*
Rasch SOS	221	84.84%	7.81%	.908

* Indicates methodology chosen for recommended estimate

Table 6
Accuracy of the RIT scale in predicting PSSA proficiency status – mathematics

Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	216	84.75%	8.14%	.904*
Second Order	216	84.75%	8.14%	.904*
Rasch SOS	216	84.75%	8.14%	.904*
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	237	89.77%	4.35%	.951*
Second Order	237	89.77%	4.35%	.951*
Rasch SOS	236	90.05%	4.84%	.946

* Indicates methodology chosen for recommended estimate

Table 7 summarizes the accuracy of prediction for this study relative to other state alignment studies. Prediction index scores for Pennsylvania are slightly higher than average in reading and a bit lower than average in mathematics. The result for mathematics was a product of a prediction index score for mathematics in grade 5 (.904) that was substantively lower than that for grade 8 (.951).

The rates of correct prediction are easily high enough to provide useful information to educators who are planning instruction to ensure all students perform at a level that meets the standards. In all grades and subjects, NWEA assessments generate a minimum of 13 correct predictions of proficiency status for each Type I error.

Table 7
Prediction Indices (Based on Proficiency Status)
for Previous NWEA State Alignment Studies

State	Reading	State	Lanaguage	State	Math
Texas	.974	Texas	.968	Texas	.970
Washington	.971	Indiana '01	.907	Wyoming	.961
Minnesota	.944	Colorado '03	.903	Colorado '01	.957
Pennsylvania	.935	Indiana '03	.894	Washington	.949
Wyoming	.931	Arizona	.874	Illinois	.946
Colorado '03	.931			California	.944
Illinois	.928			Colorado '03	.943
California*	.921			South Carolina	.943
Arizona	.912			Minnesota	.936
Colorado '01	.910			Washington	.936
Nevada	.902			Pennsylvania	.926
South Carolina	.902			Arizona	.919
Indiana '01	.902			Indiana '01	.899
Indiana '03	.900			Nevada	.866
Washington	.886			Indiana '03	.860

* California and Texas results were generated by a study of over 1,000 per grade from a single school district.

Predicting PSSA Performance Levels from RIT Scores

The PSSA reports four levels of performance. Three cut scores are set to define these four levels. Analyzing the capacity of RIT scores to predict students' PSSA performance levels can help educators triangulate information about student performance on their state test, assuring that instructional plans and interventions are adequately reinforced by data. Predictions of performance level are not as accurate as the predictions of proficiency status. This is true in part because tests vary in their ability to measure students at the highest and lowest performance levels. The *advanced* levels on the Grade 8 tests are harder to estimate with precision because so few students attained this level of performance.

When predicting performance levels, a case is identified as accurate when the performance level assigned by the PSSA and RIT score are the same. A Type I error occurs when the RIT score assigns a performance level that is higher than the student actually achieved on the state test. For example, if the RIT score projects an *advanced* performance for the student and the PSSA result is *proficient*, we declare the case a Type I error because the RIT score overestimated performance.

Table 8
Accuracy of the RIT scale in predicting PSSA performance level – reading

Grade 5	Accuracy	Type I Error	Prediction Index	% Advanced found	% Below Basic found
Linear	61.07%	21.06%	.670	48.64%	68.86%
Second Order	64.66%	15.69%	.757*	59.92%	77.41%*
Rasch SOS	64.08%	20.75%	.676	74.71%*	74.41%*
Grade 8	Accuracy	Type I Error	Prediction Index	% Advanced found	% Developing found
Linear	64.65%	17.21%	.734	20.55%	68.86%
Second Order	65.67%	14.14%	.785*	38.36%	79.53%*
Rasch SOS	63.26%	20.28%	.679	54.79%*	77.91%

* indicates most accurate method

Table 9
Accuracy of the RIT scale in predicting PSSA performance level – mathematics

Grade 5	Accuracy	Type I Error	Prediction Index	% Advanced found	% Below Basic found
Linear	63.86%	19.35%	.697	70.00%	66.12%
Second Order	63.86%	19.35%	.697*	70.00%	66.12%
Rasch SOS	64.08%	20.82%	.675	79.67%*	71.90%*
Grade 8	Accuracy	Type I Error	Prediction Index	% Advanced found	% Developing found
Linear	73.21%	13.40%	.817	18.18%	81.63%
Second Order	74.51%	11.35%	.848*	52.27%	87.20%
Rasch SOS	74.88%	13.30%	.822	72.73%*	85.34%*

* indicates most accurate method for this purpose

The results reported in tables 8 and 9 suggest that second order regression generally produced the best overall estimates of performance level, while the Rasch SOS was more often successful in finding the most students performing at the lowest and highest performance levels.

NWEA has reported estimated performance level assignments for prior studies conducted in 11 states. Table 10 compares the accuracy with which these tests predict performance level. The results show the PSSA performance index scores slightly below the median in both reading and mathematics.

Table 10
Prediction index scores by performance level assignment
for previous NWEA state alignment Studies

State	Reading	State	Math
Washington	.874	Washington	.928
Texas	.868	Texas	.900
Indiana	.860	Illinois	.888
Colorado	.840	Colorado	.808
Illinois	.804	Washington	.805
Nevada	.776	Indiana	.804
Pennsylvania	.770	Pennsylvania	.769
South Carolina	.757	South Carolina	.764
Arizona	.756	Arizona	.756
Washington	.698	Nevada	.742
Minnesota	.627	Minnesota	.611

Best estimates of PSSA performance level cut scores

To estimate the RIT scores that best predict the cut scores for the various Pennsylvania performance levels we did the following:

- For the *proficient* RIT score, we selected the methodology that produced the highest performance index score in predicting “pass/fail” alone.
- For the *developing/approaches* RIT score and the *advanced* RIT score, we selected the cut scores that correctly predicted the largest proportion of students who actually achieved these levels of performance on the PSSA.

Table 11 summarizes the recommended cut scores for each performance level on the PSSA. Based on NWEA student growth norms, the table also includes estimated cut scores for grades 6 and 7 that would indicate “on-track” performance for students who will be taking the grade 8 PSSA test in that subject.

Table 11
Projected RIT Scores Equivalent to Performance Levels on PSSA
 (estimated scores for years not tested are in blue)

Reading	Developing			Approaches	Proficient			Advanced		
	Score Range	% of pop. identified	Method	Cut Score	Cut Score	Perf. Index	Method	Cut Score	% of pop. Identified	Method
Grade 5	<=198	77.41%	2 nd Order Rasch	199	212	.928	2 nd Order	221	74.71%	Rasch
Grade 6	203			204	217			227		
Grade 7	207			208	220			231		
Grade 8	<=210	79.53%	2 nd Order	211	223	.944	2 nd Order	235	54.79%	Rasch
Mathematics	Developing			Approaches	Proficient			Advanced		
	Score Range	% of pop. identified	Method	Cut Score	Cut Score	Perf. Index	Method	Cut Score	% of pop. Identified	Method
Grade 5	<=205	71.90%	Rasch	206	216	.904	Linear 2 nd Order Rasch	228	79.67%	Rasch
Grade 6	211			212	222			235		
Grade 7	216			217	229			242		
Grade 8	<=222	87.20%	2 nd Order	223	237	.951	Linear 2 nd Order	250	72.73%	Rasch

Using RIT scores to estimate student probability of achieving passing performance on the PSSA

Helping students pass the state test is not the primary reason our members use NWEA assessments. We hope they are used to provide teachers information that will allow them to improve the learning of all students. Nevertheless, state test results are important and failing to do well on them can have deleterious effects on students and their schools. Because of this, we believed educators would benefit from knowing more about the probability that a student's RIT score would lead to a passing score on the PSSA. This would allow educators to more reliably identify students who will need additional resources to reach this level of performance. Equally important, however, it will allow educators to know which students are "safe" against Pennsylvania standards so they can focus their time with these students on providing new challenges that better suit their current needs.

Tables 12 and 13 show the proportion of students at each RIT level who earned scores at or above the *proficient* level on the PSSA reading and mathematics assessments. Using Table 12 as an example, we find that about 18% of the 5th grade students who achieved a reading RIT score between 200 and 204 went on to achieve a passing score on the PSSA reading assessment. A 5th grade teacher with ten students performing in this range would know that only about two in ten of these students will be proficient on the PSSA unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 89% of 5th grade students performing at 220 to 224 level achieved proficiency on the Pennsylvania reading assessment. Teachers should feel free to focus their efforts with these students on new and more difficult challenges than the basic third grade standards might provide.

Figures 2 and 3 are graphic depictions of the data in the tables.

Table 12
Proportion of students passing the PSSA based on same spring RIT score - Reading

RIT	Grade 3	Grade 5
190	0.00%	2.13%
195	8.06%	0.00%
200	18.47%	1.33%
205	24.86%	5.22%
210	53.93%	16.42%
215	80.43%	27.27%
220	88.72%	49.25%
225	98.82%	79.63%
230	100.00%	95.12%
235		100.00%

Figure 2 - Proportion of students receiving proficient or better on Pennsylvania assessment by RIT - Reading

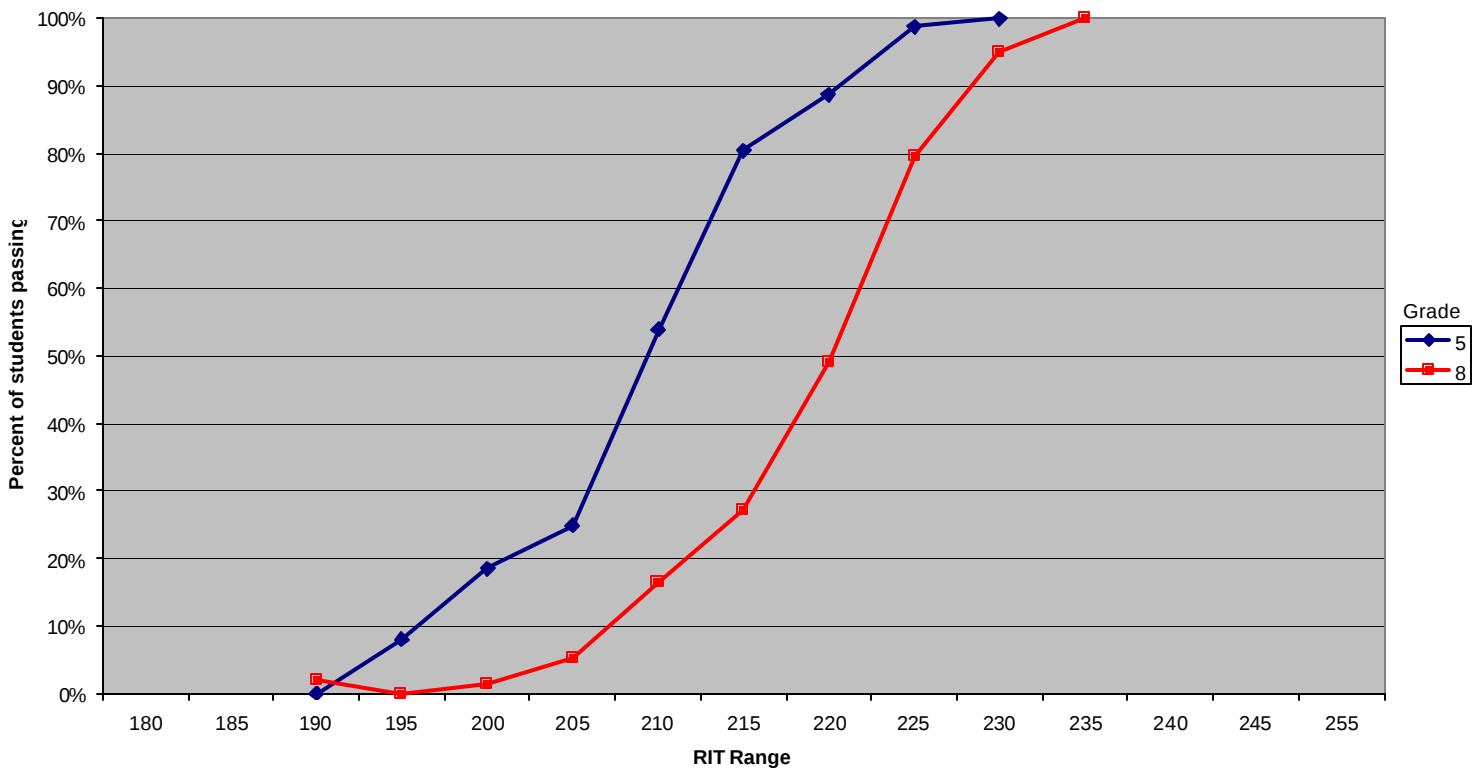
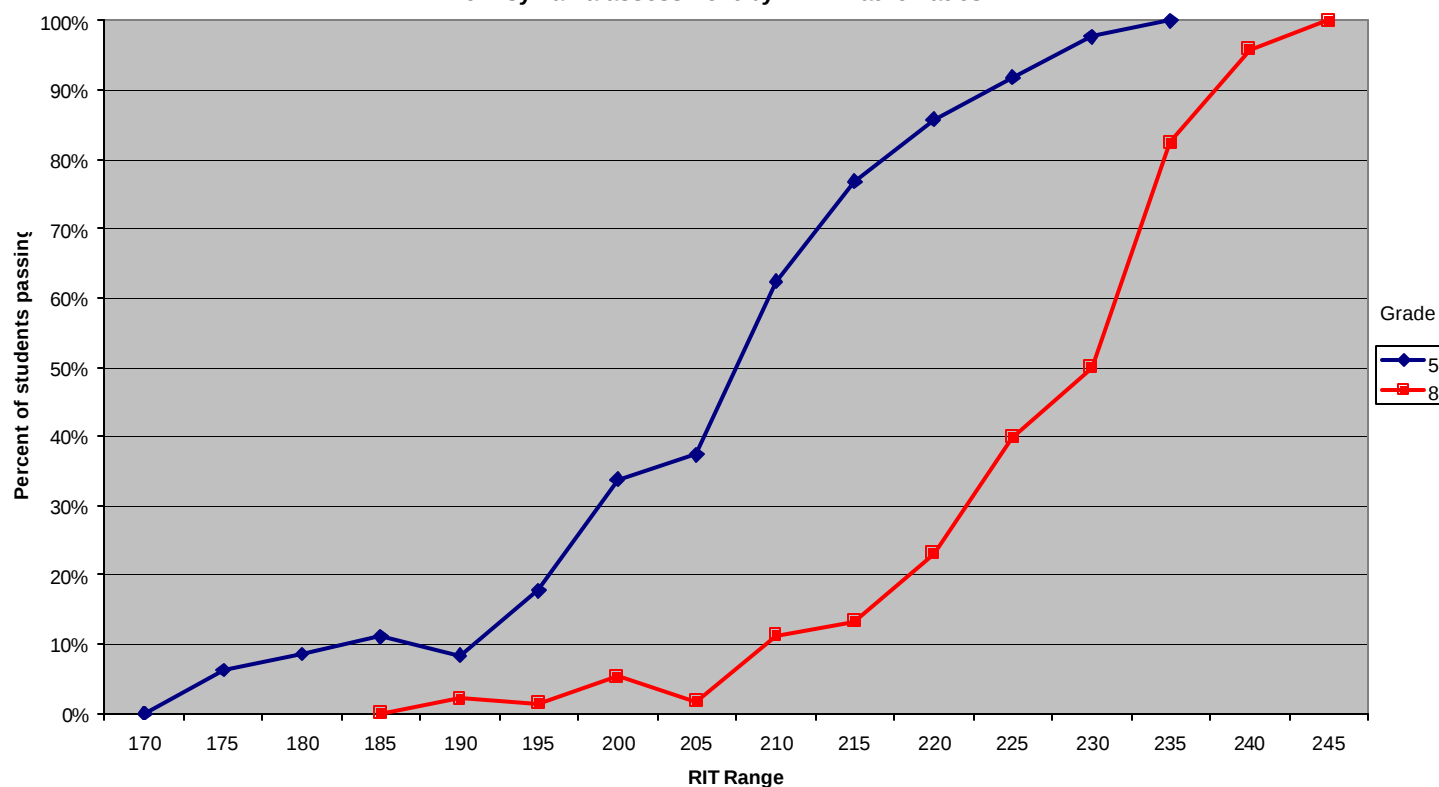


Table 13
Proportion of students passing the PSSA based on same spring RIT score - Mathematics

RIT	Grade 5	Grade 8
170	0.00%	
175	6.25%	
180	8.57%	
185	11.11%	0.00%
190	8.33%	2.13%
195	17.74%	1.45%
200	33.76%	5.33%
205	37.30%	1.74%
210	62.36%	11.19%
215	76.81%	13.22%
220	85.71%	23.13%
225	91.76%	39.81%
230	97.73%	50.00%
235	100.00%	82.46%
240		95.83%
245		100.00%

**Figure 3 - Proportion of students achieving proficient score on
Pennsylvania assessment by RIT - Mathematics**



Comparing Pennsylvania proficiency standards with the estimated standards reported in other state test alignment studies

Northwest Evaluation Association tests have been aligned with the cut scores for the state proficiency test in sixteen states. To get an estimate of the difficulty of the Pennsylvania standards in relation to other state tests, we evaluated the standard used as the cut score for *NCLB reporting or passing* and compared it to the cut score representing the same standard in these other states. Although the number of states studied is rapidly increasing, the states studied may not reflect what is typical in regard to these kinds of standards.

The results are summarized in Table 14. Pennsylvania's cut scores in both reading and mathematics are close to NWEA's national median scores in both reading and mathematics and slightly above the median of the state standards studied. We'd recommend caution about drawing any judgments about the quality of Pennsylvania's standards from that information. States establish standards for different purposes. Some states, Washington might be an example, set standards at a level they believe appropriate for students pursuing post-secondary education. Others may set standards at a lower level that reflects the literacy needed to be successful in the workplace. The No Child Left Behind Act requires schools to set targets that would result in all students achieving a *proficient standard* or proficient level of performance in about 12 years. Some communities in Pennsylvania are no doubt close to achieving this already, but many will have to improve the performance of large proportions of their students to reach this goal. Standards should be judged on how well they align with the purposes the community has set for establishing standards, not purely on how high or low the "bar" is set. One thing the tables make clear is that proficiency standards vary widely from state to state and that proficiency is not yet a concept that has a shared definition, although greater consensus in standard setting seems to be emerging. It would be fair to say that most states that we have studied who have set standards since implementation of No Child Left Behind has begun have tended to establish standards near or below the 50th percentile on our norms.

Table 14 - Cut scores representing “proficient” or “meets standards” level of performance on 16 state assessments

Reading

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	205	67	WY	214	73	SC	220	73	SC	221	63	SC	227	70	WY	232	74	MT	224	43	OR	236	77
NV	202	58	SC	213	70	NV	215	59	CA	216	46	WA	226	67	SC	230	68	IA	224	43	WA	227	53
CA	200	51	WA	207	53	CA	214	56	MT	211	35	CA	221	50	OR	227	58	ID	221	37	ID	224	44
MN	193	35	CA	205	46	PA	212	50	ID	211	35	MT	218	43	CA	226	54	CO	204	9	MT	224	44
OR	193	35	ID	200	34	AZ	210	45	IN	210	32	IA	216	37	AZ	224	49				IA	223	42
ID	193	35	MT	196	26	OR	209	42	IA	209	30	ID	215	35	PA	223	46				CO	209	15
MT	193	35	IA	196	26	IL	207	37	TX	208	28	TX	210	24	IN	219	35				CA	208	14
IL	193	35	CO	191	18	MT	206	35	CO	197	11	CO	206	18	MT	219	35						
IN	192	32				ID	206	35							IA	219	35						
IA	191	31				IA	205	32							ID	218	32						
AZ	190	29				MN	204	30							IL	218	32						
TX	179	13				TX	204	30							MN	218	32						
CO	179	13				CO	197	18							CO	206	12						

Mathematics

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	208	75	WY	221	83	SC	227	76	SC	235	78	SC	242	78	WY	257	89	MT	242	47	WA	257	73
CA	204	63	WA	218	76	CA	225	71	CA	230	68	WA	242	78	SC	251	80	IA	241	44	MT	247	40
NV	203	59	SC	217	74	AZ	220	59	IN	221	47	CA	238	71	AZ	248	75	ID	240	42	IA	247	40
IN	201	50	CA	212	59	NV	216	48	ID	219	42	ID	225	44	CA	240	60	CO	235	32	OR	245	33
OR	199	46	ID	205	39	PA	216	48	IA	218	40	MT	224	42	PA	237	53				ID	242	25
AZ	199	46	IA	205	39	OR	215	46	MT	218	40	IA	222	38	OR	235	50				CO	233	14
MN	198	42	MT	205	39	ID	213	41	CO	207	19	TX	221	35	ID	233	46				CA	232	13
MT	197	39				MT	212	38				CO	216	26	MN	231	42						
IA	197	39				IA	212	38							IN	231	42						
ID	196	36				MN	210	33							IL	230	40						
IL	193	29				IL	210	33							MT	228	36						
						TX	209	31							IA	228	36						
						CO	201	15							CO	225	31						

- Indiana tests students in the fall. Their cut scores were adjusted to reflect equivalent spring performance
- Colorado uses the partially proficient level of performance for NCLB reporting. To maintain consistency we report the level each state uses for NCLB reporting here.
 - The Texas estimate is based on the level for proficient performance that will be implemented in 2005.

Using RIT scores and data from this alignment study to set individual growth targets

NWEA encourages educators and parents to collaborate on setting individual growth targets for students based on what we call a “hybrid-growth model”. The *proficient standard* cut score for each grade reflect benchmarks that students who are “on-target” would meet if they were to achieve the state’s benchmark for the *No Child Left Behind Act*. For students who are behind this benchmark, we recommend a growth target that would reflect the norm for their grade and RIT range (see the 2002 NWEA norms study for this information) plus some proportion of the gap between their current performance and the benchmark that the student would try to close during this school year. For those students whose performance is ahead of the benchmark, we suggest a target that reflects their current RIT range norm.

This approach assures that each student has a growth target that is challenging. It also assures that low performing students have targets that will assure they eventually reach proficiency standards. Schools that achieve high rates of success on these kinds of targets will assure that no child is left behind (to borrow a phrase) while also making sure that all children have the opportunity to get ahead, regardless of where they stand against a standard. More information on this approach can be obtained by contacting the research team at NWEA.

Summary and Conclusions

This study investigated the relationship between the scales used for the PSSA assessments and the RIT scales used to report performance on Northwest Evaluation Association tests. The study determined RIT score equivalents for the PSSA performance levels in reading and mathematics. Test records for more than 2,400 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project PSSA performance levels. Second-order regression methods generally produced the most accurate cut score estimates. Accuracy of predicting PSSA passing performance was above 84% for all grades when using the best methodology. Type I errors ranged from about 4% to 8% when the best methodology was employed.

Readers should exercise some caution about generalizing these results to their own settings. Curricular or instructional differences unique to your districts may influence the accuracy with which the estimated cut scores reflect actual performance in your setting. With this limitation in mind, we would encourage educators to use this data as one tool to inform standards-based decisions.

The information gathered in this study came from measures employing the NWEA RIT Scale. Because all of the research that we have to date indicates that scores generated from computer-based tests and Achievement Level Test (ALT) scores are virtually interchangeable, readers should feel comfortable applying the results of this study in any setting that uses the RIT scale.

We hope that data from this study provides useful information to help Pennsylvania educators use NWEA assessments to better inform, plan and deliver student instruction. Good information, when matched with the professionalism and commitment of our Pennsylvania colleagues, will assure that every student has the opportunity to reach their aspirations.

References

Kingsbury, G. (1999, April) A comparison of test scores from the Iowa Test of Basic Skills and Meridian Checkpoint Assessment Level Tests. Portland, OR. Northwest Evaluation Association.

Kingsbury, G., Breithaupt, D. and Hauser, C. (2002). An empirical comparison of scores from NWEA assessments and the Iowa Test of Basic Skills. Portland, OR. Northwest Evaluation Association.